UC Irvine
UC Irvine Previously Published Works

Title
Unresolved direction of host transfer of Plasmodium vivax v. P. simium and P. malariae v. P. brasilianum

Permalink
https://escholarship.org/uc/item/2fd700qs

Journal
Infection Genetics and Evolution, 11(1)

ISSN
1567-1348

Authors
Tazi, Loubna
Ayala, Francisco J

Publication Date
2011

DOI
10.1016/j.meegid.2010.08.007

Supplemental Material
https://escholarship.org/uc/item/2fd700qs#supplemental

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Infection, Genetics and Evolution 11 (2011) 209–221

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Unresolved direction of host transfer of *Plasmodium vivax* v. *P. simium* and *P. malariae* v. *P. brasilianum*

Loubna Tazi [a,*], Francisco J. Ayala [b,**]

[a] Division of Epidemiology, Human Genetics and Environmental Sciences, University of Texas, Health Science Center at Houston, School of Public Health, Brownsville Regional Campus, Brownsville, TX 78520, USA

[b] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA

ARTICLE INFO

Article history:
Received 22 January 2010
Received in revised form 3 July 2010
Accepted 13 August 2010
Available online 16 September 2010

Keywords:
Evolutionary history
Recent bottleneck
Host switch
Plasmodium vivax
Plasmodium malariae
Polymorphisms

ABSTRACT

The evolutionary history of two human malaria parasites, *Plasmodium vivax* and *Plasmodium malariae*, remains unresolved. The near genetic identity between human *P. vivax* and *P. malariae*, and primate *P. simium* and *P. brasilianum*, respectively, suggests that recent host transfers occurred, but questions remain, such as whether the transfer was from humans to New World monkeys or vice versa, and when the transfers occurred. Here, we investigate the phylogenies, haplotype networks, positive selection and genetic diversity among these parasite species by means of four genes. Human *P. vivax* and primate *P. simium* recently derived one from the other; at least two host transfers have occurred. Human *P. malariae* and primate *P. brasilianum* also have recently derived one from the other by lateral host transfer. The direction of the host transfer cannot be decided in either one of the two pairs of species, owing to the scarcity of available strains from the primate parasites.

Published by Elsevier B.V.

1. Introduction

Malaria represents a severe problem of public health worldwide, with approximately 350–500 million clinical cases and 1–3 million deaths annually (Breman et al., 2004; Nahlen et al., 2005). Among the four species of malaria parasite that infect humans (*Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, and *Plasmodium ovale*), *P. vivax* is the most geographically widespread. It accounts for 70–80 million clinical cases annually, across much of tropical and subtropical Asia, Central and South America, the Middle East, and Africa (Mendis et al., 2001). *P. falciparum* is the most malignant species, as it accounts for 80% of human malaria's morbidity and mortality, mostly in sub-Saharan Africa.

Determining the evolutionary history of *Plasmodium* parasites and their hosts holds a great biological relevance for understanding the pathogenicity and life-history traits of primate malarias (Martinsen et al., 2008; Rich et al., 2009). To date, molecular phylogenetic studies have mainly focused on the origin of *P. falciparum* and its relationships to other human and animal malaria parasite species (Waters et al., 1991; Escalante and Ayala, 1994; Escalante et al., 1995, 1998, 2004; Joy et al., 2003; Rich and Ayala, 2003; Jeffares et al., 2007). Until recently, there has not been general agreement about the origin and age of the extant populations of *P. falciparum* (Rich and Ayala, 2000; Volkman et al., 2001; Mu et al., 2002; Joy et al., 2003; Hartl, 2004; Escalante et al., 2005; Hagner et al., 2007; Jeffares et al., 2007). Indeed, the origin of *P. falciparum* still remains controversial. However, four recent papers (Rich et al., 2009; Prugnolle et al., 2010; Krief et al., 2010; Duval et al., 2010) indicate that *P. falciparum* originated from parasites of African great apes, most likely from *P. reichenowi*, a common chimpanzee parasite.

Much less is known about the evolutionary history of *P. vivax*, perhaps in part because its lesser virulence has not stimulated as much research as for *P. falciparum*, but also because it is difficult to maintain this parasite under *in vitro* continuous culture conditions (Golenda et al., 1997). Phylogenetic analyses have placed *P. vivax* among the Asian primate malarias (Escalante et al., 1995, 2005; Mu et al., 2005; Cornejo and Escalante, 2006). Nevertheless, the high prevalence of Duffy negativity (absence of the Duffy blood-group antigen, which protects against *vivax* malaria infection) among human populations in sub-Saharan Africa has been interpreted as evidence favoring the African origin of *P. vivax* (Livingstone, 1984; Carter and Mendis, 2002; Carter, 2003).

* Corresponding author at: University of Texas, Health Science Center at Houston, Division of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, Brownsville Regional Campus, 80 Fort Brown, SPH Bldg Room S.100, Brownsville, TX 78520, USA. Tel.: +1 956 882 5163; fax: +1 956 882 5152.

** Corresponding author at: University of California, Irvine, Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, Irvine, CA 92697-2525, USA. Tel.: +1 949 824 8293; fax: +1 949 824 2474.

E-mail addresses: Loubna.Tazi@uth.tmc.edu (L. Tazi), fjayala@uci.edu (F.J. Ayala).

1567-1348/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.meegid.2010.08.007

Recent studies have shown a paucity of synonymous sites in *P. vivax*, which may suggest that the world expansion of human *vivax* malaria is of recent origin (<10,000 years ago) (Cui et al., 2003; Feng et al., 2003; Leclerc et al., 2004; Lim et al., 2005). A study, based on polymorphisms in two nuclear genes and one plastid gene, suggests that the extant populations of *P. vivax* originated 45–82 Kya (Escalante et al., 2005). The discovery of near genetic identity between *P. vivax* and *P. simium*, a platyrrhine monkey parasite, raises the possibility that a host transfer has occurred between humans and New World monkeys in very recent evolutionary times (Escalante and Ayala, 1994; Ayala et al., 1999; Leclerc et al., 2004; Escalante et al., 2005). Moreover, a study based on polymorphisms in the *Csp* (circumsporozoite protein) gene has shown that at least two host transfers have occurred between humans and New World monkeys, with alternative reasons given that favor one or the other of the two possible directions of host transfer (Lim et al., 2005).

Questions about the origin and the evolutionary history of another human malaria parasite, *P. malariae*, also remain largely unresolved, partly because it is not as predominant as it may once have been and, thus, it has failed to stimulate as much research as *P. falciparum* or even *P. vivax*. *P. malariae* is genetically indistinguishable from *P. brasilianum*, a parasite infecting New World monkeys in Central and South America, which again suggests a recent host transfer between humans and primates (Escalante et al., 1995, 1998; Qari et al., 1996). The direction of the transfer remains also unsettled, with available reasons favoring one or the other of the two possible directions of host transfer, New World monkeys to humans or vice versa (Ayala et al., 1998).

Our study combines, for the first time, the analysis of sequences available in GenBank for four different genes, three protein-coding genes: *Csp* (circumsporozoite protein), *Cytb* (cytochrome *b*), *Msp-1* (merozoite surface protein-1), as well as the *SSU rRNA* gene. These sequences have not previously been analyzed together. In contrast to previous works, our study investigates the evolutionary and phylogenetic relationships, not only between *P. vivax* and *P. simium*, but also between *P. malariae* and *P. brasilianum*. Studies focusing on this last pair of species still remain limited, if not inexistent. In our study, we did not intend to show that *P. vivax* and *P. simium* are close relatives and that *P. malariae* and *P. brasilianum* are also close relatives. Those facts have been already documented in previous studies. Our work aimed to compare the evolutionary history between these two pairs of species and to determine if the host transfers in those species occurred from humans to New World monkeys. The originality of our work also relies on the unique combination of the analytical approaches used, which would allow us to predict the direction of transfer of these pathogens between humans and New World monkeys. According to data generated from previous studies, we expect a scarcity of synonymous substitutions in both human malarial species. A recent origin of both human parasites is therefore expected, with a possible scenario that host transfers occurred from humans to monkeys after the colonization of the Americas. Our results support a recent transfer in both cases, although we cannot settle with certainty the direction of the host transfer, owing particularly to the limited availability of strains of the two primate malarias, *P. simium* and *P. brasilianum*. Our PAML analysis also indicated the lack of sites under diversifying selection for the genes studied, with the exception of the *Csp* gene in *P. vivax* and *P. simium*. This finding is consistent with the recent origin of one of these two parasite species. Moreover, no positively selected site was identified in the *Csp* and *Msp-1* genes in *P. malariae* and *P. brasilianum*. This result suggests that *P. malariae* and *P. brasilianum* may have diverged even more recently than *P. vivax* and *P. simium*. In any case, the scarcity of neutral polymorphisms suggests that both, *P. vivax* and *P. malariae*, are recent human parasites or, alternatively, that they have experienced recent population bottlenecks.

2. Materials and methods

2.1. DNA sequences

Table 1 lists the geographic origin of the malaria strains and their accession numbers for the coding sequences of the *Csp*, *Cytb* and *Msp-1* genes, as well as the *SSU rRNA* gene sequences. The genes of *P. vivax*, *P. malariae*, *P. simium*, and *P. brasilianum* are from isolates collected from global malarial endemic regions. The number of sequences varies greatly among species and genes, with fairly numerous strains of *P. vivax Csp* (24 sequences), *Cytb* (32 sequences), and *SSU rRNA* (8 sequences); but only 1–4 strains of *P. malariae Csp*, *Msp-1*, and *SSU rRNA*; and only one, two, or three of *P. simium* and *P. brasilianum* for any given gene. All sequences were obtained from GenBank and represent the sequences available in this database for each species and the respective genes. In comparison with the *P. vivax* sequences, there is unfortunately a significant scarcity of sequences in the other species for the genes analyzed. Sequences from *P. cynomolgi*, *P. simiovale*, *P. knowlesi*, and *P. gonderi*, also from GenBank, are used for outgroup comparisons. Our total dataset consists of 95 sequences. The *Csp* and *Msp-1* genes include a large central region of repeats that vary in number and sequence. The *Csp* sequences, but not the *Msp-1* sequences, from GenBank included the central repeat region. The *Csp* central repeat region is only taken into account in the phylogenetic reconstruction of the strains, but not in the other analyses.

2.2. Alignment and phylogenetic analyses

The sequences for each gene were aligned using MUSCLE, a multiple sequence alignment software (Edgar, 2004), and the resulting alignments were checked using MacClade 4.05 (Maddison and Maddison, 2000). In order to get more reliable alignments for all genes in our phylogenetic analyses, Gblocks v0.91b was used to select the conserved blocks in the alignments (Castresana, 2000; Talavera and Castresana, 2007). The phylogenetic analyses were performed using Maximum Likelihood (Felsenstein, 1981) with nodal support assessed via bootstrapping (1000 pseudoreplicates) (Felsenstein, 1985) as implemented in PhyML (Guindon and Gascuel, 2003); and Bayesian methods coupled with Markov Chain Monte Carlo (BMCMC) inference as implemented in MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003). Model selection for these analyses followed the procedure outlined by Posada and Buckley (2004) as implemented in ModelTest v3.6 (Posada and Crandall, 1998). For the BMCMC techniques, two independent analyses were run, each consisting of four chains. Each Markov chain started from a random tree and run for 2.0×10^7 cycles, sampling every 1000th generation. In order to confirm that our Bayesian analyses converged and mixed well, we monitored the fluctuating value of likelihood and compared means and variances of all likelihood parameters and likelihood scores from the independent runs using the program Tracer v1.4.1 (Rambaut and Drummond, 2009). All sample points prior to reaching stationarity were discarded as burn-in. The independent runs gave identical results for each of the four genes.

2.3. Haplotype network reconstruction

A haplotype network was constructed for each gene using the method of statistical parsimony (Templeton et al., 1992), as implemented in the software package TCS v1.21 (Clement et al., 2000). Due to the genetic relatedness between *P. vivax* and *P. simium*, and between *P. malariae* and *P. brasilianum*, the haplotype

Table 1
Plasmodium species, geographic origin, haplotype, accession numbers, and VK types for all analyzed sequences of four genes (*Csp*, *Msp-1*, *Cytb*, and *SSU rRNA*). Primate parasites are in boldface type.

Gene	Species	Geographic origin	Haplotype	Accession number	VK type
Csp	*P. vivax*	Solomon Islands	H1	U08983	210
	P. vivax	Mauritania	H2	AY674050	210
	P. vivax	South Korea	H3	AF164604	210
	P. vivax	South Korea	H4	AF164606	210
	P. vivax	South Korea	H5	AJ292977	210
	P. vivax	South Korea	H5	AJ278611	210
	P. vivax	South Korea	H5	AJ400910	210
	P. vivax	North Korea	H5	M20670	210
	P. vivax	China	H5	U08977	210
	P. vivax	China	H5	U08978	210
	P. vivax	China	H5	U08979	210
	P. vivax	China	H6	S73385	210
	P. vivax	Philippines	H7	U08980	210
	P. vivax	Philippines	H8	U08981	210
	P. vivax	Gabon	H8	U09737[**]	247/210/210
	P. vivax	Brazil	H9	M11926	210
	P. vivax	Solomon Islands	H10	U08982	210
	P. vivax	Bangladesh	H11	AY843440[**]	247/247/210
	P. vivax	Thailand	H11	M34697	210
	P. vivax	El Salvador	H11	J02751	210
	P. simium	**Brazil**	**H11**	**L05068**	**210**
	P. vivax	Papua New Guinea	H12	M69060	247
	P. vivax	Brazil	H12	M69061	247
	P. vivax	Brazil	H12	M69062	247
	P. simium	**Brazil**	**H12**	**L05069**	**247**
	P. vivax	Papua New Guinea	H13	M69059	247
	P. brasilianum	**unknown origin**	**H14**	**J03203**	**247**
	P. malariae	China	H15	S69014	247
	P. malariae	Ivory Coast	H16	AJ001526	247
	P. malariae	Uganda	H16	J03992	247
	P. malariae	Cameroon	H17	AJ001523	247
	P. cynomolgi	**unknown origin**		**M15102**	
	P. simiovale	**unknown origin**		**U09765**	
	P. knowlesi	**unknown origin**		**M11031**	
Msp-1	**P. brasilianum**	French Guiana	H18	AF138878	
	P. brasilianum	French Guiana	H19	AF138882	
	P. malariae	French Guiana	H20	AF138881	
	P. cynomolgi	**unknown origin**		**AY869723**	
	P. knowlesi	**unknown origin**		**X91855**	
Cytb	*P. vivax*	Vietnam	H21	AY598127	
	P. vivax	India	H21	AY598120	
Cytb	*P. vivax*	Solomon Islands	H21	AY598118	
	P. vivax	Bangladesh	H21	AY598115	
	P. vivax	Brazil	H21	AY598097	
	P. vivax	Thailand	H21	AY598073	
	P. vivax	Pakistan	H21	AY791629	
	P. vivax	Sri Lanka	H21	AY791575	
	P. vivax	East Timor	H21	AY791573	
	P. vivax	North Korea	H21	AY791554	
	P. vivax	Panama	H21	AY791548	
	P. vivax	Colombia	H21	AY791547	
	P. vivax	Honduras	H21	AY791545	
	P. vivax	Dominican Republic	H21	AY791544	
	P. vivax	El Salvador	H21	AY791543	
	P. vivax	Nicaragua	H21	AY791541	
	P. vivax	Mauritania	H21	AY791527	
	P. vivax	Brazil	H21	AY791537	
	P. simium	**unknown origin**	**H21**	**NC_007233**	
	P. simium	**unknown origin**	**H21**	**AY722798**	
	P. simium	**South America**	**H21**	**AY800110**	
	P. vivax	Ethiopia	H22	AY791525	
	P. vivax	Papua New Guinea	H23	AY598119	
	P. vivax	Iran	H23	AY791626	
	P. vivax	Vanuatu	H23	AY791574	
	P. vivax	Philippines	H23	AY791571	
	P. vivax	Borneo	H23	AY791555	
	P. vivax	Indonesia	H24	AY598108	
	P. vivax	Melanesia	H25	AY791692	
	P. vivax	Papua New Guinea	H26	AY791656	

Table 1 (*Continued*)

Gene	Species	Geographic origin	Haplotype	Accession number	VK type
	P. vivax	Namibia	H27	AY791528	
	P. vivax	Vanuatu	H27	AY791561	
	P. vivax	China	H27	AY791600	
	P. vivax	China	H27	AY598139	
	P. vivax	El Salvador	H28	NC_007243	
	P. cynomolgi	**Malaysia**	**H28**	**AY800108**	
	P. simiovale	**Malaysia**	**H28**	**AY800109**	
	P. knowlesi	**unknown origin**	**H28**	**AY722797**	
	P. gonderi	**Africa**	**H28**	**AY800111**	
SSU rRNA	*P. vivax*	Thailand	H29	U93235	
	P. vivax	El Salvador	H30	U93095	
	P. vivax	Brazil	H31	AY579418	
	P. vivax	unknown origin	H32	U83877	
	P. vivax	unknown origin	H33	X13926	
	P. simium	**unknown origin**	**H34**	**AY579415**	
	P. simium	**unknown origin**	**H35**	**U69605**	
	P. vivax	Papua New Guinea	H36	AF145335	
	P. vivax	El Salvador	H37	U03080	
	P. vivax	El Salvador	H38	U07368	
	P. brasilianum	**French Guiana**	**H39**	**AF130735**	
	P. malariae	Greece	H40	AF014942	
	P. malariae	Thailand	H41	U78741	
	P. malariae	Papua New Guinea	H42	AF145336	
	P. malariae	Uganda	H43	M54897	
	P. cynomolgi	unknown origin		L08242	
SSU rRNA	**P. knowlesi**	unknown origin		U83876	

** U09737 has the VK247 type in the 5′-terminal region, but the VK210 type in the repetitive central region and the 3′-terminal region. AY843440 has the VK247 type in the 5′-terminal region and the central repetitive region, but the VK210 type in the 3′-terminal region.

network reconstruction (TCS network) provides indeed very useful information beyond the phylogenetic analyses. In addition to geographic information, the TCS network indicates the effects of recombination and the presence of homoplasy in the dataset. Moreover, this analysis enables us to determine the number of mutational steps between strains. The *Csp* central repeat region is not taken into account for this analysis and for the remaining analyses. Only the silent nucleotide sites in the *Csp* and *Msp-1* sequences were considered in the haplotype network reconstructions, because these two genes are subject to strong selective pressure and, therefore, amino-acid replacements may not reflect accurate evolutionary relationships between *P. vivax* and *P. simium*, or between *P. malariae* and *P. brasilianum* (due, e.g., to convergence). For this program, sequences are collapsed into haplotypes, and frequencies of the haplotypes are calculated. An absolute distance matrix is also calculated for all pairwise comparisons of haplotypes. Therefore, the networks for *P. vivax* + *P. simium* and *P. malariae* + *P. brasilianum* are constructed so that the squares (haplotypes with the highest outgroup probability) and circles (other haplotypes) represent actual sequences derived from the strains analyzed. The size of the squares and circles is proportional to the number of sequences displaying the same genotype. Each small solid circle represents a putative sequence in the evolutionary pathway. The solid lines on a network represent one-step mutational connections among distinct genotypes (observed or putative) with at least a 95% degree of confidence, whereas the dashed lines represent a more tenuous connection. The networks for all sequences analyzed in the different malaria species provide no evidence for recombination events and homoplasy. All our analyses are therefore not compromised.

2.4. Tests of selection

We used the likelihood methods implemented in the CODEML program in the PAML package v4 (Yang, 2007) for detecting selection acting on specific codons in the protein-coding genes (*Csp*, *Msp-1*, and *Cytb*). We estimated ω (dN/dS ratio) and the proportion of sites (p) with $\omega > 1$ using the codon-based nested models M1 (neutral), M2 (selection), M7 (beta), and M8 (beta and ω) (Goldman and Yang, 1994; Yang et al., 2000). Model likelihood scores were compared using a Likelihood Ratio Test (LRT) to determine the best-fit model. When two models are nested (M1 vs. M2, M7 vs. M8), twice the log-likelihood difference was compared with a χ^2 distribution where the degrees of freedom equal the difference in the number of free parameters between the two models. When ω is greater than 1 in M2 or M8, positively selected sites are inferred. We used the empirical Bayesian approach to identify the potential sites under diversifying selection as indicated by a posterior probability ($pP > 0.95$) (Nielsen and Yang, 1998; Yang et al., 2005).

2.5. Genetic diversity estimates

The genetic diversity (θ) was determined using LAMARC v2.0.2 (Kuhner, 2006), which provides a Maximum Likelihood estimate of theta ($\theta = 2N_e\mu$; N_e is the effective population size and μ is the mutation rate per site) for comparing the genetic diversity between *P. vivax* and *P. simium*, and between *P. malariae* and *P. brasilianum*. Three independent runs were performed for this analysis to ascertain the reproducibility of the genetic diversity estimates. In addition to theta, we also estimated other polymorphism statistics between strains and between the two pairs of species of interest, including π_S, the average pairwise differences per site between sequences, as well as the number of segregating synonymous sites *S*, using the program DnaSP v5 (Librado and Rozas, 2009). For the coding sequences (*Csp*, *Msp-1* and *Cytb*), only the silent nucleotide sites were considered. In our study, we used different programs in order to strengthen the conclusions drawn from the genetic diversity estimates.

Fig. 1. Phylogenetic relationships among strains of *Plasmodium vivax* and *Plasmodium simium* based on three genes: *Csp* (a), *Cytb* (b), and *SSU rRNA* (c). The arrows highlight the position of *P. simium* strains in the phylogenetic trees. The *Csp* sequences are of two types, VK210 and VK247, which differ by three diagnostic amino-acid replacements in the two terminal regions of the gene, as well as by the repeated amino-acid "motifs" that make up the central region (Lim et al., 2005). The VK type is noted for each strain (the strain numbers for 247 are in boldface type). Notice that the *P. simium* strains are one of each type. Two *P. vivax* strains are recombinant (marked with **; see Table 1). Numbers above branches indicate bootstrap support (ML methods) and numbers below branches indicate Bayesian posterior probabilities converted to percentages.

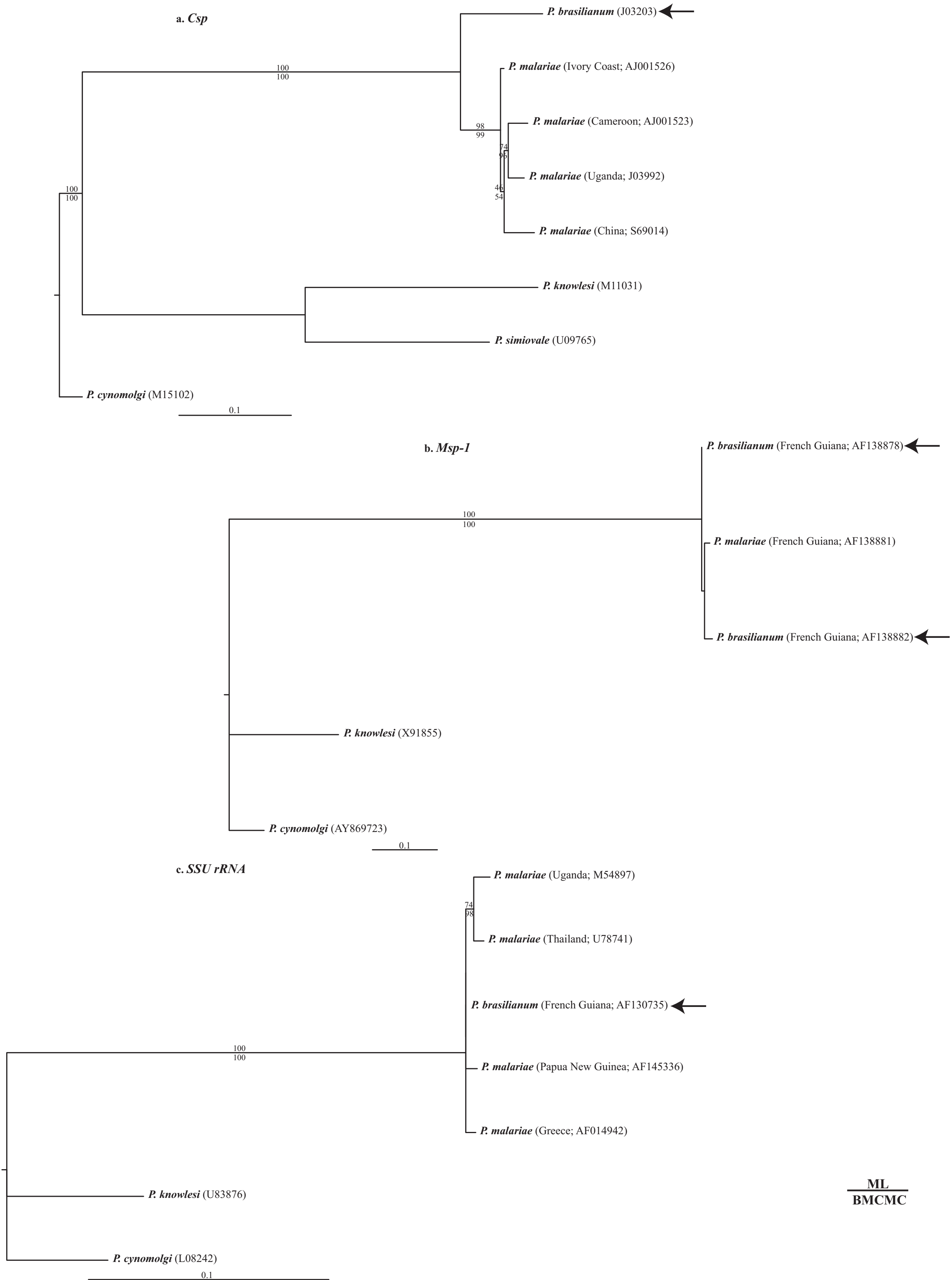

Fig. 2. Phylogenetic relationships among strains of *Plasmodium malariae* and *Plasmodium brasilianum* based on three genes: *Csp* (a), *Msp-1* (b), and *SSU rRNA* (c). The arrows highlight the position of *P. brasilianum* strains in the phylogenetic trees. Numbers above branches indicate bootstrap support (ML methods) and numbers below branches indicate Bayesian posterior probabilities converted to percentages.

3. Results

Fig. 1 displays the phylogenetic relationships among the *P. vivax* and *P. simium* strains for each of the three genes studied in these two species (*Csp*, *Cytb*, and *SSU rRNA*). Fig. 2 shows the phylogenies derived from the *Csp*, *Msp-1*, and *SSU rRNA* genes among the *P. malariae* and *P. brasilianum* strains. The outgroups yield similar phylogenetic trees for all the genes analyzed, confirming indeed the stability of the topology of these trees. Notice, however, the position of one *P. simium* strain (U69605) in the *SSU rRNA* phylogeny (see Fig. 1c), which is closer to *P. cynomolgi* than to *P. vivax* or to the other *P. simium* strain (AY579415). For the other two genes, the *P. simium* strains fit well within the *P. vivax* clade, and far removed from *P. cynomolgi* and the other two species.

Overall, the phylogenetic analyses confirm the similarity between *P. vivax* and *P. simium*, as well as between *P. malariae* and *P. brasilianum*, with the major clades showing good statistical support, represented by high bootstrap values and posterior probabilities (see Figs. 1 and 2). The only exception, as just noted, is observed in the phylogeny derived from the *SSU rRNA* gene, where a strain of *P. simium* (U69605) is clustered with *P. cynomolgi* rather than with *P. vivax* or the other *P. simium* strain (AY579415) (see Fig. 1c). Both Maximum Likelihood and Bayesian approaches support this observation with significant statistical support. Indeed, the genetic distance (number of nucleotide substitution) between the U69605 *P. simium* strain and *P. cynomolgi* strain is 0.067, smaller than the genetic distance between the two *P. simium* strains (0.090) (see Table 2), and also smaller than the average distance between the two *P. simium* strains and all the *P. vivax* strains (0.082 ± 0.007), or between the two *P. simium* strains and the combination of all *P. vivax* and *P. cynomolgi* strains (0.082 ± 0.007), although these differences lack statistical significance. Notice that the differences in average distances of the two *simium* strains to all *vivax*, or to all *vivax* plus *cynomolgi* and/or *knowlesi*, or to these two other species combined, are not statistically significant (see Table 3).

Measures of genetic diversity (θ), in addition to other polymorphism statistics (S and π_S), are shown in Table 4. Because of the small number of *P. simium* strains (two or three per locus), the genetic diversity (θ) values for this species have large confidence intervals. This analysis indicates greater genetic diversity in *P. vivax* than *P. simium* at the two protein coded locus, but not at *SSU rRNA*, where the two *simium* strains are considerably different from one another, as noted above and resulting in the phylogenetic position of one *simium* strain close to *P. cynomolgi* (see Fig. 1c). In any case, the values of θ and π_S for *P. vivax* are not significantly different from those of *P. vivax* + *P. simium* (see Table 4). In the *Cytb* gene, $\pi_S = 3.7 \times 10^{-4}$ in *P. vivax* strains, whereas no polymorphic site ($\pi_S = 0$) has been identified among the three *P. simium* strains. In the *Csp* gene (the central repeat region of which is not considered for estimating the genetic parameters in Table 4), $\pi_S = 1.0 \times 10^{-3}$ for *P. vivax* strains, but no polymorphic site occurs in *P. simium* strains ($\pi_S = 0$). The θ and π_S values of the *P. malariae* strains are quite similar to the combination of *P. malariae* + *P. brasilianum* at any of the three loci investigated in these two species, but the confidence intervals of θ are very large due to the small number of strains (see Table 4).

The number of segregating synonymous sites in *P. vivax* and *P. simium* is low, with only three among the 24 strains of *P. vivax* in the *Csp* gene and only two among the 32 strains of the same species in the *Cytb* gene (see Table 4). There are no segregating synonymous sites in *P. simium* in either gene. In contrast, the number of segregating synonymous sites is relatively high in *P. malariae* and *P. brasilianum* in the *Csp* and *Msp-1* genes, respectively. In fact, the analysis of the *Csp* gene in *P. malariae* shows two segregating synonymous sites among the four strains studied; and two segregating synonymous sites among the two strains of *P. brasilianum* are detected in the *Msp-1* gene (see Table 4). In the *SSU rRNA* gene, there is a high number of polymorphic sites in *P. vivax* (183) and *P. simium* (68), but few (only six) in *P. malariae* and *P. brasilianum* (see Table 4).

Haplotype networks for *P. vivax* and *P. simium* using *Csp*, *Cytb*, and *SSU rRNA* genes are shown in Fig. 3; and for *P. malariae* and *P. brasilianum*, using *Csp*, *Msp-1*, and *SSU rRNA* genes, are shown in Fig. 4. The *Csp* and *Msp-1* haplotypes do not include the central repeat region; only silent nucleotide sites are considered in these genes. The *P. vivax*–*P. simium* haplotype networks are considerably more complex than those of *P. malariae*–*P. brasilianum*, with several unique genotypes relative to the number of sequences analyzed, doubtless owing to the larger number of strains of the first pair of species. There is some geographic clustering, but a

Table 2
Genetic distance between *SSU rRNA* sequences from *Plasmodium vivax*, *Plasmodium simium*, *Plasmodium cynomolgi*, and *Plasmodium knowlesi*, calculated using the PAUP* v4.0b10 software (Swofford, 2002).

	1*vi*	2*vi*	3*vi*	4*vi*	5*vi*	6*vi*	7*vi*	8*vi*	9*si*	10*si*	11*cy*	12*kn*
1 *P. vivax* (U83877)		.002	.002	.060	.085	.085	.118	.117	.014	.081	.097	.116
2 *P. vivax* (X13926)			.000	.058	.083	.083	.117	.115	.012	.080	.095	.114
3 *P. vivax* (AY579418)				.058	.083	.083	.117	.115	.012	.080	.095	.114
4 *P. vivax* (AF145335)					.039	.039	.131	.129	.072	.090	.070	.105
5 *P. vivax* (U07368)						.003	.150	.148	.094	.071	.048	.085
6 *P. vivax* (U03080)							.150	.148	.094	.071	.048	.085
7 *P. vivax* (U93095)								.002	.131	.136	.159	.183
8 *P. vivax* (U93235)									.129	.134	.157	.181
9 *P. simium* (AY579415)										.090	.106	.125
10 *P. simium* (U69605)											.067	.102
11 *P. cynomolgi*												.088
12 *P. knowlesi*												

Table 3
SSU rRNA genetic distances (with standard deviation) between various combinations of species and the two *Plasmodium simium* strains. These analyses were performed using the program MEGA v4.1 (Tamura et al., 2007; Kumar et al., 2008).

	P. simium (U69605)	*P. simium* (AY579415)	Both *P. simium*
All *P. vivax*	0.114 ± 0.009	0.070 ± 0.007	0.082 ± 0.007
All *P. vivax* + *P. cynomolgi*	0.107 ± 0.009	0.073 ± 0.007	0.082 ± 0.007
All *P. vivax* + *P. cynomolgi* + *P. knowlesi*	0.107 ± 0.008	0.078 ± 0.008	0.085 ± 0.008
P. cynomolgi	0.061 ± 0.008	0.112 ± 0.013	0.090 ± 0.010
P. cynomolgi + *P. knowlesi*	0.086 ± 0.008	0.120 ± 0.013	0.102 ± 0.010

Table 4
Polymorphism statistics (θ, π_S and S) for the four genes analyzed (*Csp*, *Cytb*, *Msp-1*, and *SSU rRNA*). Only the silent nucleotide sites were considered for the coding sequences. π_S and S were estimated using the DnaSP v5 software (Librado and Rozas, 2009). LAMARC's confidence intervals for θ are also given.

Gene	Species	n	S	π_S	θ	95% C.I.
Csp	*P. vivax*	24	3	1.0×10^{-3}	4.6×10^{-3}	1.9×10^{-3} to 9.7×10^{-3}
	P. simium	2	0	0	8.1×10^{-9}	1.0×10^{-14} to 5.0×10^{-7}
	P. vivax and *P. simium*	26	3	9.2×10^{-4}	4.4×10^{-3}	2.0×10^{-3} to 9.0×10^{-3}
	P. malariae	4	2	1.5×10^{-3}	1.9×10^{-3}	2.4×10^{-4} to 1.1×10^{-2}
	P. malariae and *P. brasilianum*	5	2	2.1×10^{-3}	1.8×10^{-3}	2.4×10^{-4} to 9.6×10^{-3}
Cytb	*P. vivax*	32	2	3.7×10^{-4}	4.1×10^{-4}	1.2×10^{-4} to 9.0×10^{-4}
	P. simium	3	0	0	1.1×10^{-6}	5.0×10^{-8} to 8.4×10^{-5}
	P. vivax and *P. simium*	35	2	3.4×10^{-4}	4.2×10^{-4}	1.8×10^{-4} to 1.0×10^{-3}
Msp-1	*P. brasilianum*	2	2	8.7×10^{-3}	8.8×10^{-3}	8.3×10^{-4} to 18.7×10^{-2}
	P. malariae and *P. brasilianum*	3	4	1.1×10^{-2}	1.3×10^{-2}	2.5×10^{-3} to 9.7×10^{-2}
SSU rRNA	*P. vivax*	8	183	9.3×10^{-2}	8.0×10^{-2}	4.1×10^{-2} to 18.8×10^{-2}
	P. simium	2	68	9.7×10^{-2}	10.7×10^{-2}	2.4×10^{-2} to 188.9×10^{-2}
	P. vivax and *P. simium*	10	137	8.2×10^{-2}	11.4×10^{-2}	6.2×10^{-2} to 23.9×10^{-2}
	P. malariae	4	6	1.3×10^{-2}	2.0×10^{-2}	6.5×10^{-3} to 8.6×10^{-2}
	P. malariae and *P. brasilianum*	5	6	1.1×10^{-2}	1.4×10^{-2}	4.9×10^{-3} to 5.0×10^{-2}

n = sample size; S = segregating synonymous sites; π_S = average pairwise differences per site; and θ = genetic diversity.

Table 5
Log-likelihood values and parameter estimates (ω, p, and n) for three protein-coding genes: *Csp*, *Msp-1* and *Cytb*.

Gene	Species	$\ln L_{M1}$	$\ln L_{M2}$	ω_{M2}	p_{M2}	n_{M2}	$\ln L_{M7}$	$\ln L_{M8}$	ω_{M8}	p_{M8}	n_{M8}
Csp	*P. vivax*	−685.2	−670.9	39.1	0.04	4	−685.4	−671	40.3	0.04	4
	P. simium	−800.1	−795.5	∞*	0.03	1	−800.1	−795.5	∞*	0.03	3
	P. malariae	−882.6	−882.6	1	0	0	−882.6	−882.6	1	0	0
Cytb	*P. vivax*	−1373.1	−1373.1	1	0	0	−1373.1	−1373.1	1	0	0
	P. simium	−1340.9	−1340.9	1	0	0	−1340.9	−1340.9	1	0	0
Msp-1	*P. brasilianum*	−313.5	−313.5	1	0	0	−313.5	−313.5	1	0	0

ln L = Likelihood values; ω = acceptance rate per site (ω_{M2} and ω_{M8}); p = proportion of sites under diversifying selection (p_{M2} and p_{M8}); and n = number of positively selected sites with a posterior probability > 0.95 (n_{M2} and n_{M8}).
* The value of ω (dN/dS ratio) in the *Csp* gene in *P. simium* ($\omega = \infty$) indicates the absence of synonymous substitutions (dS = 0) in the sequences analyzed. Only three nonsynonymous substitutions (dN = 3) are detected in these sequences.

single haplotype may include strains from very different geographic origins. This observation is consistent with a recent world expansion of *P. vivax*, with little time for geographic divergence, even between distant regions, such as Asia, America, and Africa (e.g., haplotype H21; but see also H8, H11, H12, and H27; see Table 1 and Fig. 3b). The networks display high genetic diversity among *P. vivax* strains. In the *Csp* gene, there are nine haplotypes (out of a total of 13) that consist each of a single *P. vivax* strain. All *SSU rRNA* haplotypes (H29 to H38) are represented each by only one strain, all separated by several mutations. The haplotype networks for *P. malariae* and *P. brasilianum* confirm the results showing these two species as genetically quite similar. The *Csp* haplotype network shows that the three African *P. malariae* strains (H16 and H17) are more closely related to each other than to the other strain isolated in East Asia (S69014; H15), which differs from *P. brasilianum* (H14) by only one mutational step (see Fig. 4a). Notice however that *P. brasilianum* is mainly represented as a putative outgroup displaying the highest outgroup probability in these networks (see Fig. 4).

Likelihood values and parameter estimates for different codon-substitution models of selection in the protein-coding genes are shown in Table 5. In our PAML analysis, we used two different nested models (M1/M2 and M7/M8) to test for positive selection in the three protein-coding genes (*Csp*, *Msp-1* and *Cytb*). Both models gave identical results, which excludes the effect of recombination in this analysis. The Likelihood Ratio Tests (LRTs) comparing models M2 and M8 with M1 and M7, respectively, do not reject the null model (''no selection'') in favor of the models allowing for positively selected sites, except for the *Csp* gene in *P. vivax* and *P. simium*. All models tested provide consistent and significant evidence for the lack of sites under positive diversifying selection in *P. malariae* (*Csp* gene) and *P. brasilianum* (*Msp-1* gene). Similarly, no positively selected site is detected in the *Cytb* gene in *P. vivax* and *P. simium*. The Bayesian approach identifies four and one positively selected sites in the *Csp* gene in *P. vivax* and *P. simium*, respectively, under model M2; and four and three potential sites under positive selection in *P. vivax* and *P. simium*, respectively, under model M8. All sites in model M2 are included in those of model M8 (pP > 0.95) (see Table 5; Supplementary Table S1).

4. Discussion

The circumsporozoite protein (CSP) has been extensively studied in human *Plasmodium* parasites because of its immune significance as the major surface protein of the sporozoites. The *Plasmodium Csp* gene, which encodes CSP, consists of a 5′-NR terminal region, a central region of tandem repeats, and a 3′-NR terminal region. Based on an analysis of 492 bp from the two terminal regions, Escalante et al. (1995) showed that the human parasite *P. vivax* is genetically indistinguishable from the New World monkeys' parasite, *P. simium*, which suggests a recent lateral host switch between humans and platyrrhine monkeys, although the direction of the switch (humans to monkeys, or vice versa) could not be determined on the basis of their genetic data.

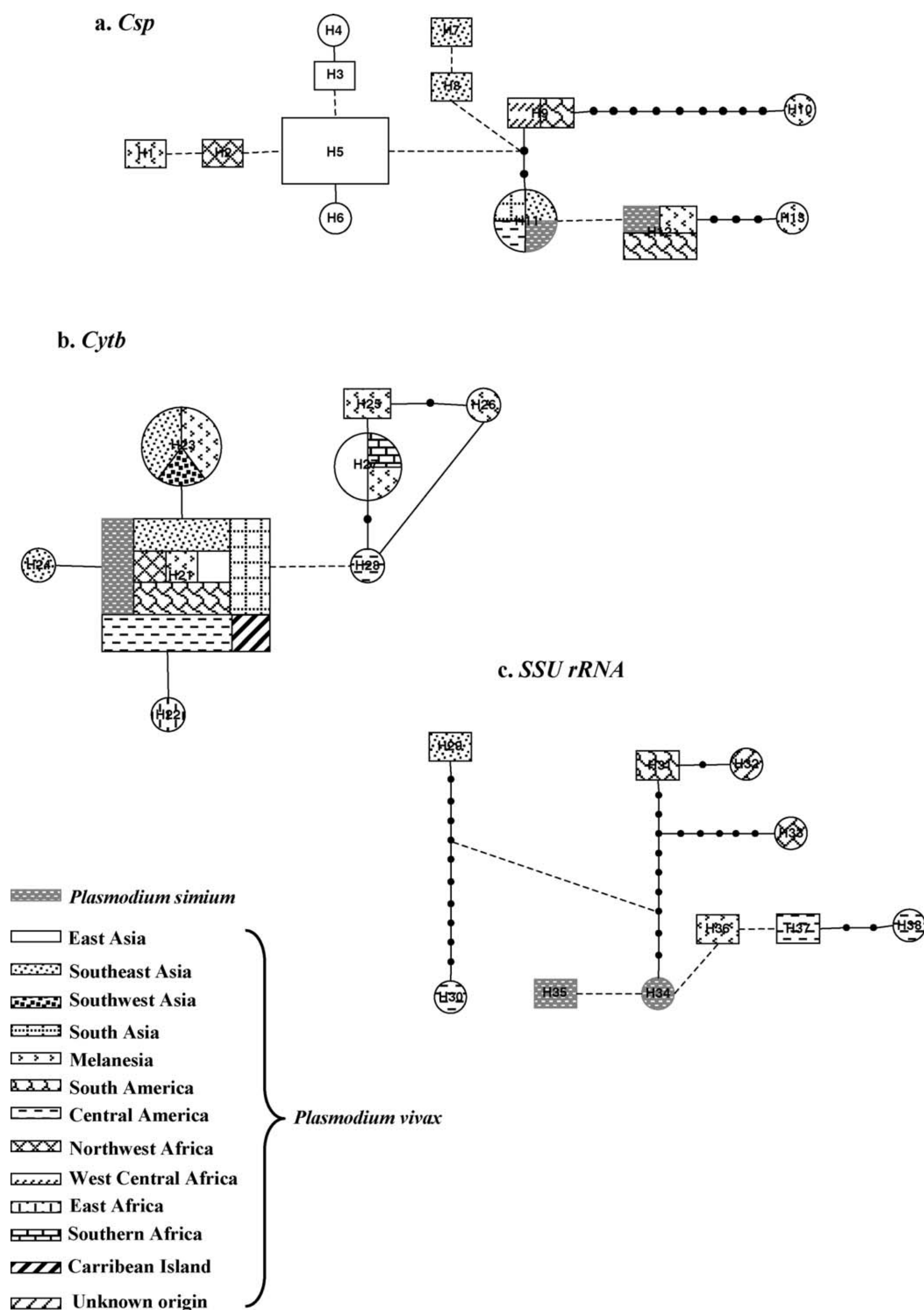

Fig. 3. Haplotype networks showing the genetic relationships among strains of *Plasmodium vivax* and *Plasmodium simium* based on the genes *Csp* (a), *Cytb* (b) and *SSU rRNA* (c). Each line segment between two circles or squares represents one mutational step. Circle and square sizes are proportional to haplotype frequencies, with each filling pattern indicating the geographic origin of the isolates. The filling patterns used for *P. vivax* and *P. simium* strains are white and gray, respectively. The small solid circles represent hypothetical intermediate haplotypes. The dash lines represent hypothetical connections between haplotypes.

Lim et al. (2005) extended the analysis to the complete *Csp* sequence and proposed that the host switch occurred at least twice. Two CSP types were earlier distinguished in *P. vivax*, VK210 and VK247, with distinctive composition of the repeating units in the central region of the protein sequence (Rosenberg et al., 1989; Kain et al., 1992; Gonzalez et al., 2001). Lim et al. (2005) found that the two CSP types, VK210 and VK247, are in fact different in each of the four regions that make up the gene: the 5′-NR and 3′-NR terminal regions, the central repeat (CR) region, and an insertion region (IR) between CR and 3′-NR. The two types, VK210 and VK247, differ by three diagnostic amino-acid replacements, one in each of the 5′- and 3′-terminal regions (5′-NR and 3′-NR) of the gene and in the insertion (IR) sequence that precedes the 3′-NR region. The central region (CR) of the gene consists of about 38 repetitive “motifs,” which are alternatively four and five amino-acids long, which also are diagnostically different between the VK210 and VK247 types. Lim et al. (2005) excluded as unlikely the possibility that the two types of sequences could have independently arisen in humans and platyrrhines by natural selection. Escalante et al. (1995) found very little genetic variation among the four strains of *P. vivax* and two of *P.*

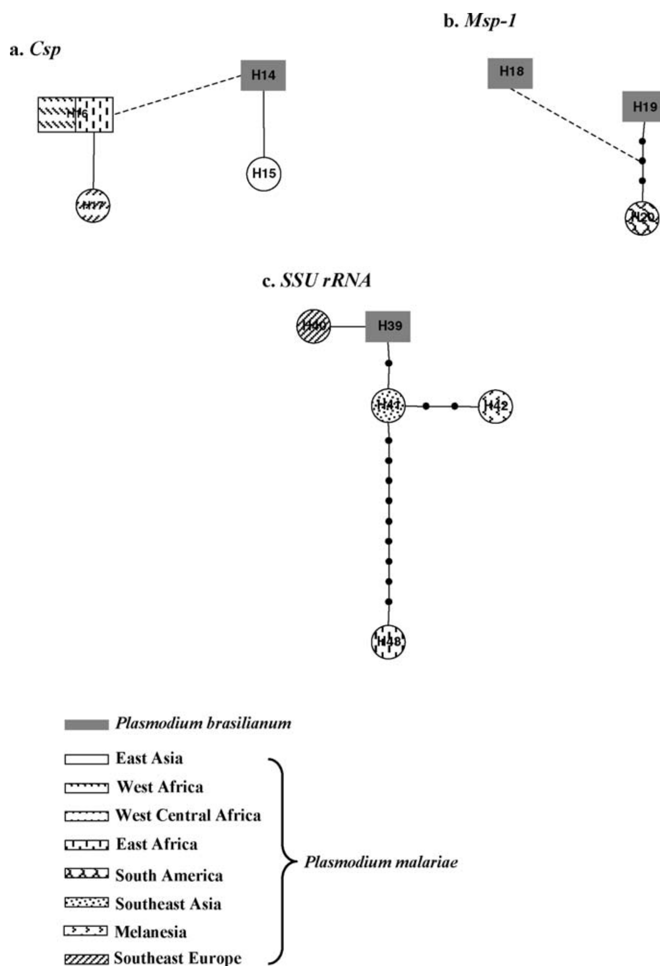

Fig. 4. Haplotype networks showing the genetic relationships between strains of *Plasmodium malariae* and *Plasmodium brasilianum* based on the genes *Csp* (a), *Msp-1* (b) and *SSU rRNA* (c) genes. Each line segment between two circles or squares represents one mutational step. Circle and square sizes are proportional to haplotype frequencies, with each filling pattern indicating the geographic origin of the isolates. The filling patterns used for *P. malariae* and *P. brasilianum* strains are white and gray, respectively. The small solid circles represent hypothetical haplotypes. The dash lines represent hypothetical connections between haplotypes.

simium in their study. Similarly, Lim et al. (2005) found very little synonymous variation among the 26 strains of both species (24 strains of *P. vivax*, representatives of the world distribution of the species, and two of *P. simium*, both from Brazil): only four segregating sites in the 5′ and 3′ regions (480 bp long combined). Amino-acid replacements were considerably more numerous: 17 polymorphic sites among the 26 strains. Amino-acid replacements may come about by natural selection. Silent substitutions, however, are likely to be neutral, or nearly so, and thus silent polymorphisms accumulate more or less gradually over time. The scarcity of silent polymorphisms among the geographically widely-representative strains of *P. vivax* was interpreted by Lim et al. (2005), as well as by Escalante et al. (1995), as evidence that the world expansion of *P. vivax* has happened in recent times, perhaps only a few thousand years ago. The lack of genetic differentiation between *P. vivax* and *P. simium* was, accordingly, interpreted as indication that the two host switches evidenced by their data have also occurred in recent times. Indeed, if the host switches occurred from humans to platyrrhines, they could not likely have happened before the human colonization of the Americas, some 15 thousand years ago, and perhaps only after the European colonizations of the last five centuries, when sedentary human populations would have made the transmission of malaria more probable.

We have analyzed the genetic polymorphism of *P. vivax* and *P. simium* in two nuclear genes, *Csp* and *SSU rRNA* and one mitochondrial gene, *Cytb*. The scarcity of silent polymorphism in the two protein-encoding genes indicates, once again, a recent world expansion of *P. vivax*. All three genes evince that *P. simium* is genetically no more different from *P. vivax* than different strains of *P. vivax* are from each other, which is, again, consistent with a recent host transfer from humans to New World monkeys or vice versa. The phylogenies displayed in Fig. 1, particularly those of *Csp* and *Cytb*, show that there is no greater genetic differentiation between strains from remote geographic regions (say, Brazil, Thailand, Gabon, and Salomon Islands—see top of Fig. 1a) than between those geographically more proximal, which is consistent, once more, with a recent world expansion of *P. vivax*.

An issue we sought to elucidate is the direction of the host switch. There are a number of biological and historical reasons why the transfer from humans to New World monkeys seems more likely than from monkeys to humans. Other reasons would seem, rather, to favor a transfer from New World monkeys to humans (Escalante et al., 1995; Ayala et al., 1998; Mu et al., 2005). If the host switch between *P. vivax* and *P. simium* was from humans to monkeys, a biological puzzle arises. Humans are biologically (evolutionarily) more closely related to Old World monkeys (catarrhines) than to New World monkeys (platyrrhines). If lateral host switch from humans to monkeys were likely, it would be more likely that the natural transfer would have been to our closer, rather than to our more remote relatives. Moreover, humans and their ancestors have been geographically associated with catarrhine monkeys for millions of years, but only for several thousand years with platyrrhine monkeys. If the natural transfer from humans to monkeys were likely, it would have been much more likely that the transfer would have occurred to species with which humans have been in geographic association for the much longer period. In a recent study, Garamszegi (2009) investigated the association between the evolutionary history of primate malaria parasites and that of their hosts, as well as the evolutionary constraints of host specialization. The concordance between host and parasite phylogenies is found to be heterogeneous: while the evolution of some malaria pathogens is strongly dependent on the phylogenetic history of their primate hosts, the congruent evolution is less emphasized for other parasite lineages (e.g., for human malaria parasites). Garamszegi (2009) suggested that the host transfer could have occurred from humans to monkeys, or vice versa. The emergence of new malaria disease in primates including humans cannot therefore be predicted from the phylogeny of parasites. In our study, both phylogeny and geographic proximity were used for predicting host transfers between humans and monkeys, as presented by Davies and Pedersen (2008).

This biological puzzle has been solved very recently, first for *P. falciparum* and *P. reichenowi* (Rich et al., 2009). Rich et al. (2009) gave a strong support for *P. falciparum* evolving from *P. reichenowi*. Varki and collaborators (Martin et al., 2005; Varki, 2007; Varki and Nelson, 2007) have also proposed that ancestors in the human lineage became resistant to *P. reichenowi* (or a related species ancestral to both *P. reichenowi* and *P. falciparum*) approximately 2–3 million years ago when a mutation in the cytidine monophosphate-*N*-acetylneuraminic acid hydoxylase gene (CMAH) led to the disruption of the biosynthetic pathway of *N*-glycolylneuraminic acid (Neu5Gc) from its precursor *N*-acetylneuraminic acid (Neu5Ac) (Chou et al., 2002; Hayakawa et al., 2006). Both molecules are erythrocyte-bound sialic acid carbohydrates, but Neu5Gc is the preferred receptor of *P. reichenowi* binding. Because of the CMAH mutation, humans fail to produce Neu5Gc, and instead have Neu5Ac, to which *P. reichenowi* has been found to have several-fold lower binding efficiency, making humans resistant to infection by this parasite (Martin et al., 2005).

This molecular adaptation is not unique since other loss-of-function mutants in erythrocyte receptor glycoproteins are also known to confer resistance of humans to other *Plasmodium* species. For example, mutation in the genes encoding the erythrocyte Duffy antigens leads to the Duffy-negative phenotype, which renders these hosts impervious to *P. vivax* infection. The Duffy-negative phenotype is extremely prevalent in sub-Saharan Africa, which explains the virtual absence of *vivax* malaria from large portions of the African continent (Mason et al., 1977a,b). Varki and collaborators argue that one *P. reichenowi* strain circulating in human hosts eventually gained an advantageous mutation in the gene encoding the erythrocyte-binding-like receptor molecule EBA-175, which binds host erythrocyte sialic acid. This mutation entailed a switch from preference for binding of Neu5Gc to a preference and high binding affinity for Neu5Ac. Hence, a fundamental adaptation to the human host environment arose in the *P. falciparum* cenancestor and has persisted in its extant progeny (Rich et al., 2009). Experimental studies indicate that *P. falciparum*'s adaptation to the novel receptor (and hence the novel host), was reinforced by the fact that the presence of Neu5Gc receptors actually interferes with the binding of *P. falciparum* EBA-175 with its preferred Neu5Ac receptor, thus ensuring that the reciprocal back transfer from humans into chimpanzees or any other catarrhines would be unlikely (Martin et al., 2005; Rich et al., 2009; Varki and Gagneux, 2009). However, it is the case that New World monkeys are also unable to synthesize Neu5Ac and, instead, accumulate Neu5Ac, its precursor in the biosynthetic pathway. Thus, *P. vivax* might have been able to invade New World monkeys, but not Old World monkeys. Thus, the discovery just described might account for both, the recent world expansion of *P. vivax* and why the host switch from humans to New World monkeys was possible while the transfer to Old World monkeys was not.

Previous studies have indicated the recent expansion of *P. vivax* and *P. malariae* in human populations. The host switch between humans and monkeys might, therefore, have occurred through a single transfer associated with a bottleneck, after the European colonization of the Americas in the 16th century. The direction of the host transfer can then be resolved by comparing the genetic diversity of the human and primate parasites.

If the transfer occurred from humans to primates, the level of genetic diversity in silent and other neutral nucleotide sites is expected to be lower in the primate parasite population than in the human parasite population. However, one can assume a variety of possible scenarios. For example, repeated transfers from the donor to the recipient host. This would increase the number of silent polymorphisms in the recipient host, but it would be expected to remain below the level in the donor—unless, of course, an extreme bottleneck would have occurred in the donor's parasite, which seems unlikely. In any case, the scarcity of available strains does not allow us to test any of the alternative scenarios, so that extended speculation would seem unwarranted.

Our study suggests that *P. simium* might have derived from *P. vivax*, whereas *P. malariae* might plausibly derive from *P. brasilianum*. For the *SSU rRNA* gene, we observed that one strain of *P. simium* (U69605) was clustered with *P. cynomolgi* whereas the other simian strain (AY579415) was mixed with the other *P. vivax* isolates. This can suggest, like a previous study (Lim et al., 2005), that the host transfer between humans and New World monkeys could have occurred twice, with the plausible scenario that *P. vivax* has been introduced first into New World monkeys following the arrival of Europeans to the New World. Moreover, the time of divergence of *P. vivax* from *P. cynomolgi*, a closely related parasite of Old World monkeys, has been estimated at 2–3 million years ago (Carter and Mendis, 2002). Considering that New World and Old World simians diverged from each other around 40–50 million years ago (Kumar and Hedges, 1998) and that they have been geographically isolated from each other, we would therefore expect that *P. vivax* could not likely have been derived from *P. simium*. Our study pointed out a relative high genetic diversity in *P. vivax* strains in comparison with *P. simium* strains, suggesting that *P. simium* could have originated from *P. vivax*. However, an increase in the number of primate strains analyzed is needed to confirm this suggestion on the direction of the transfer. For the pair *P. malariae*–*P. brasilianum*, our analyses indicated that *P. brasilianum* strains seem to be likely more divergent than *P. malariae*. *P. malariae* may have appeared therefore as a species derived from *P. brasilianum*. In addition to our general expectation, additional factors may indeed alter the expectation such as several transfer episodes between the human parasite population and the primate parasite population, severe bottlenecks in the original host parasite population, or continuous migration between the two populations after the host transfer. However, also in this case, the lack of enough sequences from both malarial species resulted in a potential limitation for drawing strong predictions for the direction of transfer between these two species.

In our study, we used sequences available in GenBank for four different genes, three protein-coding genes: *Csp*, *Cytb*, *Msp-1*, and the *SSU rRNA* gene. Those sequences were appropriate and not too conserved for testing the predictions of the host transfer in this study. Even though the overall dataset was limited in terms of number of sequences analyzed in each strain, its quality, length of sequences and power to test the predictions are very adequate. Using four different genes, instead of only one gene increases the significance of our tests. All genes analyzed showed concordant results. In addition to the genes used for this study, the SNP (Single Nucleotide Polymorphism) analysis of the genome of the different primate and human parasite species would have been another good genetic marker for the predictions of the host transfer.

Escalante et al. (1995) showed that *P. malariae* and the platyrrhine *P. brasilianum* were genetically indistinguishable with respect to the *Csp* gene. Our results corroborate that result and extend it to two other nuclear genes, *SSU rRNA* and the protein-encoding *Msp-1*. Our results indicated that *P. malariae*, similarly to *P. vivax*, has very low silent variation and, thus, that its worldwide distribution—represented in our study by strains from Asia, Africa, and South America—may be of recent origin or may have recently experienced a population bottleneck. The very limited number of strains studied makes, however, these conclusions tentative at best.

Whether the host transfer between *P. malariae* and *P. brasilianum* was from humans to New World monkeys or vice versa, elicits the same issues, and the same likely answers, as the question of the direction of the host switch between *P. vivax* and *P. simium*. The isolation of additional strains of *P. brasilianum* might, as in the case of *P. simium*, make possible to obtain the requisite genetic information for settling the matter.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2010.08.007.

References

Ayala, F.J., Escalante, A.A., Lal, A.A., Rich, S.M., 1998. Evolutionary relationships of human malaria parasites. In: Sherman, I.W. (Ed.), Malaria: Parasite Biology, Pathogenesis and Protection. ASM Press, Washington, DC, pp. 285–300.

Ayala, F.J., Escalante, A.A., Rich, S.M., 1999. Evolution of *Plasmodium* and the recent origin of the world populations of *Plasmodium falciparum*. Parassitologia 41, 55–68.

Breman, J.G., Alilio, M.S., Mills, A., 2004. Conquering the intolerable burden of malaria: what's new, what's needed: a summary. Am. J. Trop. Med. Hyg. 71, 1–15.

Carter, R., 2003. Speculations on the origins of *Plasmodium vivax* malaria. Trends Parasitol. 19, 214–219.

Carter, R., Mendis, K.N., 2002. Evolutionary and historical aspects of the burden of malaria. Clin. Microbiol. Rev. 15, 564–594.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Paabo, S., Satta, Y., Takahata, N., Varki, A., 2002. Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. Proc. Natl. Acad. Sci. U.S.A. 99, 11736–11741.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1659.

Cornejo, O.E., Escalante, A.A., 2006. The origin and age of *Plasmodium vivax*. Trends Parasitol. 22, 558–563.

Cui, L., Escalante, A.A., Imwong, M., Snounou, G., 2003. The genetic diversity of *Plasmodium vivax* populations. Trends Parasitol. 19, 220–226.

Davies, T.J., Pedersen, A.B., 2008. Phylogeny and geography predict pathogen community similarity in wold primates and humans. Proc. R. Soc. B 275, 1695–1701.

Duval, L., Fourment, M., Nerrienet, E., Rousset, D., Sadeuh Mba, S.A., Goodman, S.M., Andriaholinirina, N.V., Randrianarivelojosia, M., Paul, R.E., Robert, V., Ayala, F.J., Ariey, F., 2010. African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the *Laverania* subgenus. Proc. Natl. Acad. Sci. U.S.A. 107, 10561–10566.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Escalante, A.A., Ayala, F.J., 1994. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. Proc. Natl. Acad. Sci. U.S.A. 91, 11373–11377.

Escalante, A.A., Barrio, E., Ayala, F.J., 1995. Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. Mol. Biol. Evol. 12, 616–626.

Escalante, A.A., Cornejo, O.E., Freeland, D.E., Poe, A.C., Durrego, E., Collins, W.E., Lal, A.A., 2005. A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. Proc. Natl. Acad. Sci. U.S.A. 102, 1980–1985.

Escalante, A.A., Cornejo, O.E., Rojas, A., Udhayakumar, V., Lal, A.A., 2004. Assessing the effect of natural selection in malaria parasites. Trends Parasitol. 20, 388–395.

Escalante, A.A., Freeland, D.E., Collins, W.E., Lal, A.A., 1998. The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. Proc. Natl. Acad. Sci. U.S.A. 95, 8124–8129.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Feng, X., Carlton, J.M., Joy, D.A., Mu, J., Furuya, T., Suh, B.B., Wang, Y., Barnwell, J.W., Su, X.Z., 2003. Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. Proc. Natl. Acad. Sci. U.S.A. 100, 8502–8507.

Garamszegi, L.Z., 2009. Patterns of co-speciation and host switching in primate malaria parasites. Malar. J. 8, 110.

Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11, 725–736.

Golenda, C.F., Li, J., Rosenberg, R., 1997. Continuous in vitro propagation of the malaria parasite *Plasmodium vivax*. Proc. Natl. Acad. Sci. U.S.A. 94, 6786–6791.

Gonzalez, J.M., Hurtado, S., Arevalo-Herrera, M., Herrera, S., 2001. Variants of the *Plasmodium vivax* circumsporozoite protein (VK210 and VK247) in Colombian isolates. Mem. Inst. Oswaldo Cruz 96, 709–712.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hagner, S.C., Misof, B., Maier, W.A., Kampen, H., 2007. Bayesian analysis of new and old malaria parasite DNA sequence data demonstrates the need for more phylogenetic signal to clarify the descent of *Plasmodium falciparum*. Parasitol. Res. 101, 493–503.

Hartl, D.L., 2004. The origin of malaria: mixed messages from genetic diversity. Nat. Rev. Microbiol. 2, 15–22.

Hayakawa, T., Aki, I., Varki, A., Satta, Y., Takahata, N., 2006. Fixation of the human-specific CMP-N-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. Genetics 172, 1139–1146.

Jeffares, D.C., Pain, A., Berry, A., Cox, A.V., Stalker, J., Ingle, C.E., Thomas, A., Quail, M.A., Siebenthall, K., Uhlemann, A.C., Kyes, S., Krishna, S., Newbold, C., Dermitzakis, E.T., Berriman, M., 2007. Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. Nat. Genet. 39, 120–125.

Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., Beerli, P., Su, X.Z., 2003. Early origin and recent expansion of *Plasmodium falciparum*. Science 300, 318–321.

Kain, K.C., Brown, A.E., Webster, H.K., Wirtz, R.A., Keystone, J.S., Rodriguez, M.H., Kinahan, J., Rowland, M., Lanar, D.E., 1992. Circumsporozoite genotyping of global isolates of *Plasmodium vivax* from dried blood specimens. J. Clin. Microbiol. 30, 1863–1866.

Krief, S., Escalante, A.A., Pacheco, M.A., Mugisha, L., André, C., Halbwax, M., Fischer, A., Krief, J.M., Kasenene, J.M., Crandfield, M., Cornejo, O.E., Chavatte, J.M., Lin, C., Letourneur, F., Grüner, A.C., McCutchan, T.F., Rénia, L., Snounou, G., 2010. On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from Bonobos. PLoS Pathog. 6, e1000765.

Kuhner, M.K., 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22, 768–770.

Kumar, S., Dudley, J., Nei, M., Tamura, K., 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9, 299–306.

Kumar, S., Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. Nature 392, 917–920.

Leclerc, M.C., Durand, P., Gauthier, C., Patot, S., Billotte, N., Menegon, M., Severini, C., Ayala, F.J., Renaud, F., 2004. Meager genetic variability of the human malaria agent *Plasmodium vivax*. Proc. Natl. Acad. Sci. U.S.A. 101, 14455–14460.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Lim, C.S., Tazi, L., Ayala, F.J., 2005. *Plasmodium vivax*: recent world expansion and genetic identity to *Plasmodium simium*. Proc. Natl. Acad. Sci. U.S.A. 102, 15523–15528.

Livingstone, F.B., 1984. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. Hum. Biol. 56, 413–425.

Maddison, D.R., Maddison, W.P., 2000. MacClade 4: Analysis of Phylogeny and Character Evolution, version 4.0. Sinauer Associates.

Martin, M.J., Rayner, J.C., Gagneux, P., Barnwell, J.W., Varki, A., 2005. Evolution of human–chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolylneuraminic acid. Proc. Natl. Acad. Sci. U.S.A. 102, 12819–12824.

Martinsen, E.S., Perkins, S.L., Schall, J.J., 2008. A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. Mol. Phylogenet. Evol. 47, 261–273.

Mason, S.J., Aikawa, M., Shiroishi, T., Miller, L.H., 1977a. Further evidence for the parasitic origin of the surface coat on Malaria merozoites. Am. J. Trop. Med. Hyg. 26, 195–197.

Mason, S.J., Miller, L.H., Shiroishi, T., Dvorak, J.A., McGinniss, M.H., 1977b. The Duffy blood group determinants: their role in the susceptibility of human and animal erythrocytes to *Plasmodium knowlesi* malaria. Br. J. Haematol. 36, 327–335.

Mendis, K., Sina, B.J., Marchesini, P., Carter, R., 2001. The neglected burden of *Plasmodium vivax* malaria. Am. J. Trop. Med. Hyg. 64, 97–106.

Mu, J., Duan, J., Makova, K.D., Joy, D.A., Huynh, C.Q., Branch, O.H., Li, W.H., Su, X.Z., 2002. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. Nature 418, 323–326.

Mu, J., Joy, D.A., Duan, J., Huang, Y., Carlton, J., Walker, J., Barnwell, J., Beerli, P., Charleston, M.A., Pybus, O.G., Su, X.Z., 2005. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. Mol. Biol. Evol. 22, 1686–1693.

Nahlen, B.L., Korenromp, E.L., Miller, J.M., Shibuya, K., 2005. Malaria risk: estimating clinical episodes of malaria. Nature 437, E3 discussion E4–5.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Prugnolle, F., Durand, P., Neel, C., Ollomo, B., Ayala, F.J., Arnathau, C., Etienne, L., Mpoudi-Ngole, E., Nkoghe, D., Leroy, E., Delaporte, E., Peeters, M., Renaud, F., 2010. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 107, 1458–1463.

Qari, S.H., Shi, Y.P., Pieniazek, N.J., Collins, W.E., Lal, A.A., 1996. Phylogenetic relationship among the malaria parasites based on small subunit rRNA gene sequences: monophyletic nature of the human malaria parasite, *Plasmodium falciparum*. Mol. Phylogenet. Evol. 6, 157–165.

Rambaut, A., Drummond, A.J., 2009. Tracer: MCMC Trace Analysis Tool, 1.4.1 edition. Institute of Evolutionary Biology, Edinburgh. , Available fromhttp://tree.bio.ed.ac.uk/software/tracer.

Rich, S.M., Ayala, F.J., 2000. Population structure and recent evolution of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 97, 6994–7001.

Rich, S.M., Ayala, F.J., 2003. Progress in malaria research: the case for phylogenetics. Adv. Parasitol. 54, 255–280.

Rich, S.M., Leendertz, F.H., Xu, G., LeBreton, M., Djoko, C.F., Aminake, M.N., Takang, E.E., Diffo, J.L., Pike, B.L., Rosenthal, B.M., Formenty, P., Boesch, C., Ayala, F.J., Wolfe, N.D., 2009. The origin of malignant malaria. Proc. Natl. Acad. Sci. U.S.A. 106, 14902–14907.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rosenberg, R., Wirtz, R.A., Lanar, D.E., Sattabongkot, J., Hall, T., Waters, A.P., Prasittisuk, C., 1989. Circumsporozoite protein heterogeneity in the human malaria parasite *Plasmodium vivax*. Science 245, 973–976.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis using Parsimony, v4.0b10. Illinois Natural History Survey, Champaign.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software, version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Templeton, A.R., Crandall, K.A., Sing, C.F., 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132, 619–633.

Varki, A., 2007. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature 446, 1023–1029.

Varki, A., Gagneux, P., 2009. Human-specific evolution of sialic acid targets: explaining the malignant malaria mystery? Proc. Natl. Acad. Sci. U.S.A. 106, 14739–14740.

Varki, A., Nelson, D., 2007. Genomic differences between humans and chimpanzees. Ann. Rev. Anthropol. 36, 191–209.

Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F., Hartl, D.L., 2001. Recent origin of *Plasmodium falciparum* from a single progenitor. Science 293, 482–484.

Waters, A.P., Higgins, D.G., McCutchan, T.F., 1991. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. Proc. Natl. Acad. Sci. U.S.A. 88, 3140–3144.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.

Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118.